SOME ASPECTS OF THE SELECTION OF CANDIDATES
FOR AIRSHIP PILOT TRAINING,
U. S. NAVY

SOME ASPECTS OF THE SELECTION OF CANDIDATES
FOR AVIATION TRAINING,
U. S. NAVY

A Thesis
Presented in Partial Fulfillment of the Requirements
for the Degree Master of Science
in Public Administration

by

William Arthur Cockrell, B.Sc.

THE OHIO STATE UNIVERSITY
1940

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The selection process is a central point of interest in any personnel program and a corner stone of the entire personnel structure. Unless it is soundly conceived there can be little hope of building a first-rate-organization. Conversely, a well conceived and properly executed selection program will go a long way toward promoting effective utilization of the available manpower, especially in time of total warfare. The more specialized an activity is, the greater is the need for selection. Modern warfare is a highly specialized activity in which personnel efficiency can be gained only by devising means of finding specialized men to do the specialized jobs.[1]

The goal of selection procedures in the armed forces should be to provide means which will make possible the most effective use of the manpower available.[2] Such effective use of manpower not only implies that every man is placed in the job where he can make his greatest contribution, but also that the optimum placement of each individual is accomplished in the shortest possible time. Given sufficient time, even with inadequate methods of selection, men would probably gravitate to billets for which they are qualified. In the early stages of an emergency however, this time element is of utmost importance, and the saving in time brought about by making proper

1 Naval Leadership - Book II. Washington, D. C. U.S.Govt. Printing Office, 1948, p. 55.

2 Stuit, Dewey B. et al. Personnal Research and Test Development in the Bureau of Naval Personnel. Princeton, N. J.: Princeton University Press 1947, p. 434.

original selections and assignments may increase considerably the manpower available for combat.

Selection procedures are sometimes opposed on the grounds that all men are pretty much alike and that the average person may be made into almost anything you want by giving him the right kind of experience. This attitude denies the basic and well established fact of individual differences. Every man, having his own pattern of abilities, aptitudes and traits, does differ from every other man. This basic fact is one of the justifications for selection. Another is found in the differences that exist among jobs. Jobs differ among themselves probably as much as do individuals, with each job demanding its particular sort of muscular movements, of sensory keenness, of perception, of coordination, of endurance, and of intelligence. If we grant that jobs differ in their demands on individuals, and that individuals differ in the potentialities they bring to the jobs, selection not only becomes desirable, but is a necessity for efficient and economic operation.

Selection cannot occur unless there is a group from which selection may be made. In other words, if there is one job and only one candidate, one cannot select, one must merely appoint. If the candidates outnumber the jobs to be filled, one can screen the applicants and select the best ones available, and the more applicants there are, the more particular you can be about the ones you select for the job. Conversely, the higher the standards, the larger the number of candidates one must have to fill a given number of jobs.

Once one has decided on selection, his selection produces results in proportion to his discrimination. With a large number of candidates and a selection procedure that is known to measure their potential usefulness on the job, selection of only the very best men obviously will result in attaining optimum performance. The AAF conducted an unusual experiment during the war to show just how this economic factor works.[3]

> In order to evaluate psychological selection procedures in a group which had not been selected on the basis of aptitude measures, an experimental group of about 1300 men were admitted to pilot training without any requirements as to either aptitude or personality. This group was tested in the standard way with the AAF Qualifying Examination and the AAF Air-Crew Classification Battery, but all men were entered into training no matter how low their scores on these tests. In the same way, the men were given a careful interview by a medical officer to determine their Adaptability Rating for Military Aeronautics, but no men were disqualified on the basis of the interview.
>
> These men were entered directly into pilot preflight training, and from there followed the usual course through primary, basic, and advanced training. They were spread through many classes and schools and were given training in the usual way, mixed in with trainees who had been screened by the standard procedures. Tests and training records were maintained for this group and were analyzed after they had completed training.
>
> The value of the qualifying examination as a preliminary screen is shown by the fact that only 45 out of 520 men who failed the examination were graduated from training, whereas the yield among those passing the examination was 211 out of 751. The effectiveness of the stanine [4] for purposes of pre-

3 Dubois, Phillip H. et al. Army Air Forces Aviation Psychology Program Research Reports - Report No. 2. Washington, D. C. U.S.Govt. Printing Office, 1947. p. 200-201.

4 Pilot stanine is a standard score on a normalized 9 point scale (based on the weighted scores of a battery of performance tests) for predicting the aptitude of students for pilot training.

diction is shown by the fact that of 150 men with pilot stanines of 1, not a single individual was graduated from advanced flying training. Only 16 out of 891 men with stanines of 2 or 3 were graduated. In contrast, of 98 men with augmented stanines of 8 or 9 only 15 were eliminated for testable reasons (flying deficiency, fear, or own request).

Thus it may be seen that for "unscreened" candidates:

a) 11.5 (520/45) entrants from the group which failed the examination were required to produce one graduate;

b) 3.5 (751/211) entrants from the group which passed the examination were required to produce one graduate; and

c) Only 1.2 (98/83) entrants from the group which received the highest rating on the test battery were required to produce one graduate. The saving in time and expense of training through proper selection is quite obvious in this instance.

Similar savings in terms of training hours and expense through the use of aptitude tests as a selective device for Naval airplane pilots have been demonstrated.[5] This analysis, based on candidates who passed the aptitude tests, and not on the total applicant population as in the previous example, still showed that the tests consistently select those groups from which only a few will fail as contrasted with the groups from which a large percentage will fail. In order to obtain a given number of successful graduates (naval aviators) more entrants were required at each successively lower aptitude level. The analysis showed that in order to obtain 1000

5 "The Predictive Value of Naval Aviation Cadet Selection Tests". Navy Department Bu Med News Letter. Vol. 8, No. 8, March 1947.

graduates, 2273 entrants at the lowest acceptable aptitude level were required, whereas only 1220 entrants at the highest aptitude level were needed to produce the same results.

In addition to saving in training time and better utilization of personnel, selection is also concerned with the more remote but more far-reaching problem of vocational adjustment. Proper selection measures which adequately discriminate between candidates lessen the probability of an individual ending up in a billet for which he is not vocationally adapted and in which he will not be satisfied. Certain billets in the armed forces, particularly the billet of aircraft pilot, have a definite appeal and glamour which attract many men who are not adapted to such work. Without adequate selection devices a certain percentage of these persons will be passed in the training stage, but never become sufficiently proficient in the operational stage to derive personal satisfaction from their contribution to the total effort. This lack of satisfaction may lead to a feeling of frustration and poor morale. An individual's feeling of success in his work with its consequent happiness and satisfaction are thus conditioned in part upon matching the requirements of the job with the individual's aptitudes, which can be best accomplished by proper selection devices.

There are numerous factors, techniques or processes used in selection such as the interview, personal history information, previous scholastic record, and the one with which this study is primarily concerned: aptitude tests. In the armed services, most of the

efforts have been devoted to the use of written mental tests for the prediction of vocational aptitude. This is due to a large extent to the fact that the tests of this type may be administered to a large group of subjects simultaneously. This is a decided advantage when a large number of candidates are being processed. The tests, moreover, may be so designed that results are quantitative and yield a wide range of scores.

To be of any value tests used in selection must possess both reliability and validity. The reliability of a test is its measure of accuracy and consistency. A test is reliable if it consistently gives the same score to a person when he is retested. The retesting may be done with the same test used originally, provided practice or memory does not markedly affect the test score. In which case a duplicate form of the test containing items similar in nature but different in actual content, should be used.[6] The problem is analagous to that of determining the reliability of some physical instrument. If we measure several objects with a steel tape, then measure the same objects again with the same steel tape, we should get practically identical results. However if we measure the objects with a cloth tape, then repeat the measurements with the cloth tape, it may well be that the two sets of measurements will differ, due to stretch in the tape. Thus we may say that the steel tape is reliable and the cloth tape unreliable. By the same token we measure

6 Tiffin, Joseph. Industrial Psychology. New York: Prentice-Hall Inc. 1947/ p. 63.

a number of people with a test on two different occasions and determine whether the relative standing of the people is the same on both occasions. If so, the test meets the requirement of reliability.

To be valid, tests should actually measure what they profess to measure. A perfectly valid test would rank candidates in precisely the same relationship to one another as they would be after a trial on the job. However such perfect validity is rarely obtained. Reasonable validity is not only possible but indispensable if the test is to have any discriminating value. Selection based upon tests which have no known validity may be little different from selection determined by the turn of a card or the color of a person's eyes. Burtt considers validity the fundamental principle of test construction and says:[7]

> The tests or other measurements to be used in selecting persons for a given occupation must be evaluated by giving them to persons whose actual ability in that occupation is known and comparing efficiency in the test with efficiency in the occupation. In other words we must not devise a test that seems plausible, trust that it will work, and start using it for employment purposes. We must first test the test. If workmen who are good in the test are good in the occupation, and those who are poor in the test are poor in the occupation, then the test is valid, while if there is no consistent relation between occupational ability and test scores the test is useless. However, this principle of testing the tests is central to the whole problem and its observance marks the difference between a scientific and an un-scientific psychological approach to personnel problems.

7 Burtt, Harold Ernest. Principles of Employment Psychology. New York: Harper and Bros. 1942. p. 5

In the initial stages of a war or an emergency expediency frequently takes precedence over logical and scientific procedures. Such was the case in early 1942 when the number of persons entering training as airship pilots was greatly expanded. Due to the relatively small number of persons previously trained in such work little thought had been given to aptitude tests as an aid in selection procedures. In addition, neither qualified technical personnel nor sufficient time were available to conduct research to devise suitable tests to predict success as airship pilots. As a result the same aptitude tests which had been developed for use in the selection of airplane pilots were used to select candidates for airship pilot training. These tests, modified from time to time, according to data obtained from the rating of airplane pilots, are still being used in the selection of airship pilots. The exigencies of the service and the relative smallness of the airship program have combined to preclude any evaluation to date of these tests as a predictor of success as airship pilots. Thus we have a battery of tests, being used as an administrative procedure for selection purposes that has never been validated in one of the fields in which it is applied.

While it is true that there are many similarities between the airship and the airplane, due to their both operating in the same medium, it may well be fallacious to treat the airship as merely another type of airplane and to judge them both by the same standards. Like the surface ship, the airship is a displacement vessel

and in many ways has more in common with it than with the airplane. The airplane is a completely mechanical contrivance deriving its lift solely from its speed and through mechanical power, while the airship derives the major portion of its lift from the displacement of air by a gas which is lighter-than-air. Thus the analogy between the airship and a water borne vessel, particularly the submarine, is more suitable than that between the airship and the airplane. The airship is in effect a displacement vessel whose normal field of operations is the lower reaches of the atmospheric ocean.[8] This difference has been highlighted for years by the generic terminology used: Airplanes being referred to as heavier-than-air craft (H.T.A.) and airships being called lighter-than-air craft (L.T.A.).

Another fundamental difference between these two types of aircraft is that of speed. The airship operates in the fifty to sixty knot speed range, whereas the airplane varies from speeds of one hundred knots to those approaching and perhaps surpassing the speed of sound.

With these fundamental differences in the two types of craft it naturally follows that there are differences in the methods of operation. Therefore the requirements for operators of the two types may well differ. A comparison of the job analyses of the two types of pilots reveals some of the more important aspects of these differences.

8 Report on the Lighter-than-Air Situation made by C. E. Rosendahl, USN (Ret.) to the National Aeronautics Committee of the American Legion. 26 August 1947. p. 26.

Reaction time is an important attribute to be considered in both cases. In L.T.A., whether as a fighter pilot or as a patrol pilot, it is of the utmost importance that the pilot have rapid, almost instantaneous reactions to various conditions. In the routine mechanics of flying this requirement need not be present to such a high degree in the case of the L.T.A. pilot. Here the pilot is more concerned with the total situation and must be able to analyze the effect of various factors such as the static condition of his ship, the trim of his ship, and the existing atmospheric conditions on his craft. This basic difference is recognized in the primary instruction procedures used for these two types of aircraft. In L.T.A. the importance of performing various maneuvers such as take-offs, landings, climbs, dives and turns in a specified manner with standard throttle and control settings is stressed at all times. In airships, the combination of the static condition of the ship, (whether statically heavy, light, or in equilibrium) the trim of the ship (whether bow heavy, bow light, or in equal trim) and the differential temperature between the gas in the envelope and the surrounding air has a considerable bearing on the technique of flying. Therefore, in L.T.A. training, the importance of (1) always being aware of the static and trim conditions of the ship, (2) obtaining the most favorable combination of these two factors, and (3) realizing that the mechanics of flying will vary under these conditions is stressed. In other words, comprehension of the total situation is more important than mere operation of the controls.

Physical coordination of the individual is required to a much higher degree in H.T.A. where the pilot must accurately blend in the control of engine, aileron, elevator and rudder properly to perform even the simplest maneuver. In L.T.A. however, the pilot is concerned with only one control at a time, having a copilot actually to operate the other control. In addition he has a mechanic and various other crew members to perform various other duties. For this reason coordination between individuals or rather, the qualities of leadership are required to a greater extent in L.T.A. While a rugged individualist may well be an excellent fighter pilot, he might be in the wrong niche as an airship pilot. The crew of an operational airship will vary from eight to fifteen men and the success of its mission will therefore depend to a large extent on the ability of the command pilot to coordinate the efforts of all the members of the crew. This same kind of coordination is also required of pilots of certain types of airplanes, such as the patrol plane and bomber. The important point, however, is that this trait should be characteristic of all airship pilots. By his actions the airship pilot must give his crew a feeling of confidence in him and in his decisions, and thus develop in them a spirit of cooperation.

CHAPTER II

HISTORICAL BACKGROUND

There is nothing new or unusual in the use of tests in personnel administration. As early as the third century B.C., Plato was suggesting the aptitude and skills essential to warriors and considering tests by which to select this group for the state.[1] Tests of one kind or another have always been used for selecting men for particular tasks, with the technique being refined as advances in the science of testing are made.

The pioneer efforts in comparing efficiency on tests with efficiency on the job were made by Munsterberg about 1911 with his study of motormen of the Boston Elevated Railway.[2] The novel feature of this work was that the tests were given to actual motormen and the test scores compared with their actual service records. The comparison showed that those with a good record and with few or no accidents made somewhat higher scores in the tests than did those motormen with a bad record of accidents. Munsterberg also gave a series of tests to girls in a school for telephone operators and compared their progress in the school with their test scores. These results indicated that there was some tendency for those with satisfactory progress in learning the work of a telephone operator to make higher scores in the test than for those with unsatisfactory progress. As pointed out by Burtt the advance made in these studies

1 Mosher, W.E. and Kingsley, J.D. Public Personnel Administration New York: Harper and Bros., 1941, p. 163.

2 Munsterberg, H. Psychology and Industrial Efficiency. Boston: Houghton Mifflin, 1913, p. 320 ff.

is fundamental.[3] Previously the tests had been standardized on anybody. Now they were standardized on persons engaged in a particular occupation, and efficiency in the tests was compared with efficiency in the occupation. This same procedure, namely testing the test, is still basic.

Shortly after this, various other psychologists began to compare test scores with occupational criteria in similar fashion. Further studies on tests for telephone operators were made by McComas.[4] Scott started his work on methods for selecting salesmen, comparing test scores with sales records.[5] Rogers published the results of his investigations on tests for typists.[6] These studies and numerous others to be found in the literature are indicative of the rapidly increasing interest in tests as an aid in selection and of the recognition of the importance of standardizing the tests in the occupation to be tested.

During the first World War tests for the selection and classification of service personnel were used by almost every nation engaged in the war.[7] In England, for example, tests were devised for

1 Scott, Walter D. Principles of Employment Psychology. New York: Harper & Bros., 1940, p. 57.

4. McComas, W.C. "Some Tests for Efficiency in Telephone Operating" J. Phil. Psych. and Scientific Method. Vol. 11, 1914, pp. 293-294.

5 Scott, W. D. "Scientific Selection of Salesmen", Advertising and Selling, 1915, Vol. 25, pp. 5-6.

6 Rogers, M. S. "Psychological Tests for Typists and Stenographers", J. Appl. Psychol. Vol. 1, 1917, pp. 268-74.

7 Viteles, Morris S. Industrial Psychology. New York: W.W.Norton and Co., Inc., 1932, p. 43.

104.22

and administered to aviation pilots, aeronautical observers, hydrophone operators and numerous other groups requiring special capacities. In both Germany and France tests were developed to aid in the selection of range finder operators, chauffeurs, pilots, and various other specialized branches of the service.

The entry of the United States into the war in 1917 brought with it the largest scale experiment in the use of selection tests that had ever been attempted up to that time. One of the greatest contributions of this period was the development of the well known Army Alpha test which was eventually administered to nearly two million men. Its uses were many and varied, not the least of which was as an aid in the selection, classification and placement of personnel. Special tests were also devised for numerous specialties, including aviators, descriptions of which will be given later.

Following World War I a new emphasis was placed upon the use of tests in personnel work, due partly to the impetus given to psychological testing during the war and partly to the general movement for economy and efficiency both in industry and in governmental organizations. The widespread use of tests in industry is indicated by a survey conducted in 1940.[8] Questionnaires requesting information concerning personnel practices in use were sent to 308 different industrial concerns. Replies from 231 companies located in 25 different states and including 47 types of business activity were received. While it is acknowledged that in all probability the results

8 Scott, W.D., et al. Personnel Management. New York: McGraw Hill Book Co., Inc. 1941, p. 519.

of this survey are somewhat biased in favor of companies with fairly well defined personnel policies, (since it is reasonable to assume that companies having little or no interest in personnel problems would not take time to complete a detailed questionnaire), it is also reasonable to assume that the replies were fairly indicative of personnel practices in use at the time. The replies covered many phases of personnel work, but of prime interest is the fact that 66% of the employers reported the use of tests in some form as an aid in selection of personnel.

The history of the use of tests in the selection of aviators parallels somewhat that of selection in industry. Flying through the air seems to us so different from the tasks ordinarily undertaken by human beings that the need for special selection procedures to decide which individuals are most capable of adaptation to the new conditions and demands of aerial flight has long been recognized. Many tests were developed during the first World War, but very few were evaluated properly. Most of them stressed general intelligence, memory, and various forms of reaction time.[9]

Reports of some of the developmental work carried on during the first World War and in the interim between wars is found in Manual on Examination and Personality Studies in Aviation Medicine[10], which states in part:

9 Flanagan, J.C., et al. Army Air Forces Aviation Psychology Research Reports - Report No. 1. Supt. of Documents, U.S.Govt. Printing Off., 1948, p. 10.

10 Technical Manual 8-320, War Dept., Washington: U.S.Govt. Printing Off., January 1941, par. 142-143.

> During the early part of the World War the Allies selected their pilots in a haphazard manner. Frequently they were assigned to the Air Corps because of their inability to continue performance of ground duties. After many terrible accidents with their tremendous toll of man-power and material, the Allies began to consider the problems connected with flying and wonder if all individuals were adequately endowed to meet these problems. In the beginning, courage was considered the only trait essential to piloting an airplane. If an individual possessed that to a high degree, there was nothing to prevent him from flying. Bitter experience taught the folly of any such assumption. No more do all individuals possess the aptitude for military flying than do all individuals possess the aptitude for painting, sculpturing, music, par golf or any of the other accomplishments to which a favored few may attain.

In Naval Aviation the selection of candidates for pilot training was initially a leisurely affair. Throughout the twenties and early thirties candidates consisted of persons already in the naval service. The number of these candidates was quite small and screening consisted of a Flight Physical Examination, involving rigorous standards of physical fitness, with special emphasis on visual standards, which was administered by the Aviation Medical Examiners. With the advent of the Naval Aviation Cadet Program in the mid-thirties, Selection Boards were located at various training bases, and functioned as part of the routine activity of these bases. The normal procedure of these boards varied somewhat from time to time and from place to place, but in general they tended to fit into a characteristic pattern. The young applicant for flight training was rapidly checked for the requirements of age, height and weight, and schooling, followed by the Flight Physical Examination. If the candidate successfully passed these hurdles he was interviewed at length by one or more line officers. According to Jenkins, while this inter-

view played a major role in determining whether the candidate was acceptable for aviation training, it was subject to enormous variability.[11] In his words:

> The interview was a tradition of Naval Aviation handed down from the last war. As a proper prerogative of line officers (by no means always pilots themselves) it was presumed to determine whether the candidate had what it took to make a Naval Aviator. With such an ambiguous precept the interview took as many forms as there were Boards; and it often found enormous variability of expression within a given Board. In some Boards, it became essentially a stress interview. In others, it served to implement the prejudices of the Commanding Officer or of the individual interviewer. In only a few was it reduced to a vestigal determination that the candidate had no obvious stigmata.

With the outbreak of war in Europe in 1939 new impetus was given to the problem of developing practical procedures for the prediction of aptitude for flying. The subsequent declaration of a State of Emergency and the advent of compulsory military training introduced several factors which made the problem of selection of pilots much more difficult. In normal times with a small flow of candidates, selection could be based on an intelligent analysis of each individual by a specially trained Flight Surgeon. This procedure was impracticable with the tremendously increased demand for pilots under emergency and war conditions. In peacetime there was also a large amount of self selection.[12] The applicants were in general those who had definite aeronautical interests and aspirations,

11 Jennings, John A. "Naval Aviation Psychology", American Psychologist. 1946, I, p. 45-49.

12 Flanagan, J.C. et al. Army Air Forces Aviation Psychology Program Research Reports - Report No. 1. Superintendent of Documents, U.S.Govt. Printing Office, 1947, p. 52 ff.

who believed that their special aptitudes fitted them especially for this type of work. Under compulsory military training other factors played a larger part in inducing men to apply for aviation training. Some individuals with little aptitude for this type of duty were attracted by the social prestige, salary, or similar expected gain, or considered it merely as a way to avoid a type of duty considered less fitting.

At this time the Civil Aeronautics Authority obtained its first allotment of funds for research on problems of the selection and training of aircraft pilots.[13] These funds were administered by a committee of the National Research Council composed of psychologists, physicians, physiologists, engineers, pilots and representatives of the military services. This group conducted research at approximately forty universities and other centers, including military establishments throughout the country. In this research program considerable attention was given to problems of pilot selection.

As Viteles states, this committee made definite contributions to Naval Aviation:[14]

> One major practical outcome of the Committee Research program is the fact that by 1941, when the United States entered the war, the research program had already produced test material and findings which were used by the U. S. Navy in setting up procedures for the selection of pilots. At the Annual Meeting of the Committee on Selection and Training of Aircraft Pilots, held in 1943, Cdr. (then

13 Viteles, Morris S., et al. The Aircraft Pilot — 5 Years of Research. National Research Council, Washington D.C. June 15, 1945, p. III.

14 Ibid., p. 15.

> Lt. Cdr.) J. G. Jenkins, in discussing the Navy research program, reported that, 'We are now using in routine selection, both before and after the beginning of training, three tests. These three tests were either developed by the Committee first of all as selection agencies in aviation, or were developed by the collaborative efforts of the Committee and the Navy.'

The three tests mentioned above were introduced in December 1941 as initial screening tests and were pencil-and-paper tests "selected on the basis of empirical evidence that they would consistently and reliably differentiate between groups who ultimately passed and ultimately failed in Naval Aviation Training."[15] (It should be noted here that the empirical evidence was obtained from N.T.S. training records exclusively, and not from A.T.A. training records.) These tests consisted of a general intelligence test, a mechanical comprehension test, and a biographical inventory.

The general intelligence test first used was the Wonderlic-Hovland Personnel Test (PT), an abridged form of the well known Otis Self-Administering Test. Three forms of this test were used, each form having fifty items and a time limit of twelve minutes. Since this test had relatively low reliability a second form was given to those applicants who made unsatisfactory scores on the first test. This procedure was designed to minimize the number of selections resulting from errors of measurement. There was also evidence of inequality of the various forms of the test.[16] These

15 "The Predictive Value of Naval Aviation Cadet Selection Tests". Navy Department, Bureau Aeros Letter, Vol. 6, No. 4, March 1947.

16 Flores, Donald M., "Validation of Naval Aviation Cadet Selection Tests Against Cadet Criteria". J. Appl. Psych. Vol. 31, No. 6, December 1947, p. 601.

limitations led to the replacement of the PT in October 1942 by the Aviation Classification Test, which was developed specifically for use in Naval Aviation. This test, which is still in use, is worded to have a Navy flavor. It contains 111 items, with a 45 minute time limit, and it deals with Practical Judgment, Arithmetic, Vocabulary, Meter-Reading and Comparisons. Two forms of this test are now in use. A list of sample problems in this test is given in Appendix I.

The Mechanical Comprehension Test (MCT) was developed by George K. Bennett of the Psychological Corporation.[17] This test, which is also in use at the present time, consists of 76 two-choice and three-choice items dealing with pictured mechanical situations, thus minimizing the verbal factor. The purpose of this test is to measure the applicant's ability to handle the mechanical concepts of everyday life. The 45 minute time limit for this test permits a large percentage of the candidates to attempt all items.

Scores on the MCT and ACT are expressed on a five-point letter scale of A, B, C, D, and E. These letters correspond to the following percentages of population: 7, 24, 38, 24, 7. To pass the MCT and ACT, the applicant must score "C" or better.

The Biographical Inventory was developed in 1940-41 by W. Lowell Kelly and others under the auspices of the CAA-NRA Committee

17 Ibid. p. 602.

on the Selection and Training of Aircraft Pilots.[18] Flight surgeons had long sought to appraise a candidate's personal history, interests and attitudes in attempting to predict his probable performance as a pilot. The BI, based upon statistical analysis of thousands of pilot records, does what the flight surgeon tried to accomplish on the basis of professional skill, experience and intuition. It is a non-time-limit questionnaire, originally containing 150 items and currently containing 169 items on biographical history, interests, habits, attitudes, and preferences. Unlike the ACT and MCT, the BI had no *a priori* right or wrong answers. Its key was based on an analysis of the responses of cadets who later passed or failed in flight training. The BI was used only as an advisory instrument until May 1944, since which time it has been used in actual selection. At the same time, and after considerable experimentation, three keys for scoring the BI were placed in use.[19] These keys were developed by item analysis of the BI for each of several ability groups as differentiated on the basis of ACT and MCT scores. One key was designed for high ability groups, the second for groups in the middle range, and the third for low ability groups.

Some objection to the use of the Biographical Inventory was

18 "The History and Development of the Biographical Inventory", AAF Division of Research, Report No. 77. October 1946, p. 12 ff.

19 "Construction of Keys A, T, and E for the Biographical Inventory". Navy Department, Aviation Psychology Technical Memorandum No. 3. 1 November 1944.

made on the grounds that it was a subjective questionnaire in which an applicant could try to outguess the Inventory by giving "good" answers instead of true ones. This objection has been answered by Jenkins.[20]

> We have two safeguards. One is that the blank as it is now used contains an approximately equal admixture of 'silent' and significant items. An applicant who wished to 'fudge' would have to guess which were the significant items and then he would have to guess which answer to give. The 'right' answers are by no means obvious, as many actual trials have shown.
>
> The second safeguard, of course, lies in the pattern aspect . . . No one item is particularly important; it is the total pattern that counts. The applicant who decided to take a wishful view of the facts on item No. 17 alters his score only to a minute degree. And experience has shown that the effort to paint a good picture almost inevitably leads him to twist certain answers negatively to compensate for others where he has succeeded in giving a positive slant to the facts.

In December of 1942 the Flight Aptitude Rating (FAR) was introduced. As the name indicates, this is intended as an index of the flight ability of the individual. This FAR was computed according to a simple formula from the combined scores on the MCT and PI. Somewhat later, when large numbers of candidates were available to fill small quotas, the Flight Aptitude Rating was used in selection — the minimum qualifying score being raised or lowered as the needs of the service dictated.

20 Jenkins, John G., "Prediction of Flight Training Performance by Biographical Data." Journal of Aviation Medicine, 1944, 15, pp. 134-35.

CHAPTER III

THE SELECTION OF SAMPLES AND VARIABLES

While the possibilities of human flight were demonstrated by the Montgolfier brothers and by Professor Charles in 1783, it was well over a century later before the simple balloon was made dirigible, or directable. The main stumbling block to the solution of this problem was an efficient power plant. The discovery of petroleum and the invention of the internal combustion engine toward the end of the nineteenth century brought the long sought key to this problem. Successful flights in large rigid airships by Count Zeppelin and in smaller non-rigid airships by Santos Dumont were both made prior to the historic flight of the Wright Brothers at Kitty Hawk in 1903. (Lighter-than-air ships are of three classes: rigid, semi-rigid, andnon-rigid. The rigid airship has a complete metal skeleton, which gives the ship strength and shape. The lifting gas is carried in several separate gas cells, nested within the bays of the ship. The semi-rigid type has a metal keel extending the length of the ship, to which control surfaces and the control car are attached, and with a metal cone to stiffen the bow section. The non-rigid airship, or blimp, has no internal support. The shape of the bag is maintained by keeping the internal gas at a higher pressure than the surrounding atmosphere.)

Airship development and flying were carried out principally by Europeans until shortly before our entry in the First World War. As a result of the successful use of airships by the British and French

during this war, the Navy entered the airship field in February 1917 when the construction of sixteen blimps was started.[1] No one in this country, however, knew much about building airships and even less was known about flying them after they were built. Flight training therefore was conducted under rather adverse conditions. Men had to teach themselves to fly airships, then teach others to fly them.[2] Most of this training was conducted at Akron, where some of the ships were built, and at Pensacola in conjunction with W.T.A. training.

At the inception of the Navy's rigid airship program after World War I, airship activities were centered at Lakehurst, N. J. The number of pilots remaining on active duty, who were trained to fly blimps during war, was not sufficient to meet the needs of the projected airship program, so a pilot training program was necessary. It was felt at that time that sufficient difference existed between the heavier-than-air and lighter-than-air branches of naval aviation to warrant entirely separate courses of training. As a result, an Airship Training School was established at the Naval Air Station, Lakehurst, New Jersey, and the first class convened in July 1923, for a twelve month course.

The size of the annual classes which attended this school between 1923 and 1940 varied according to the demand for pilots to operate the fluctuating numbers of airships in commission. Condi-

1 Allen, Hugh. The Story of the Airship (Non-Rigid). Chicago: The Goodyear Press, 1931, p. 12.

2 Ibid. p. 14.

dates for training were all members of the regular Navy, serving in other branches of the service, who were selected on the basis of their previous record and ability to pass the Flight Physical Examination. By 1940 a total of 157 officers had successfully completed the training course and had been designated Naval Aviator (Airship). Of this total, 92 officers remained on active duty in the Navy. With the contemplated expansion of the Navy and of the lighter-than-air facilities it soon became evident that the small flow of trainees would require augmentation. The best source of supply of additional trainees seemed to be the Naval Aviation Cadet Training Program. Starting with a modest quota of five cadets in 1940, trainees from this source steadily increased until the completion of the training program in 1946. In addition, an augmented number of officers and enlisted pilots were trained during this period.

The impact of the national emergency and of our entry into the war resulted in accelerated training with an initial reduction in the length of the course to six months, and finally, starting in July 1942, to four months. Also, in October 1942 the training of Airship pilots was started at Moffett Field, California.

With the formation of the Naval Airship Training Command in May 1943, pilot training at Lakehurst and Moffett Field was coordinated for the first time. Shortly thereafter a revised training program was placed in effect, whereby the course was lengthened to six months, with primary training being conducted at Moffett Field and advanced training at Lakehurst. This plan enabled the two schools

to specialize and thus improve the quality of instruction.[3]

Upon successful completion of the training and designation as Naval Aviator (Airship) cadets were commissioned and assigned to fleet airship squadrons engaged in coastal patrol and convoy operations. These squadrons operated in the Atlantic, Pacific, and Gulf coastal waters of the United States, and in the coastal waters of the Caribbean, Eastern Central America and Brazil. During the winter of 1944 one squadron made the first trans-oceanic non-rigid airship flight and operated in the Mediterranean until the end of hostilities. With the exception of a selected few retained as instructors in the training command, cadets were assigned to the squadron of their own choice whenever the needs of the service permitted.

The history of airship pilot training has been briefly reviewed in order to make clear the chain of reasoning by which the subjects for this study were chosen, and to indicate the relatively small total population involved. The total personnel trained as airship pilots from 19 October 1940 to 30 June 1944 is summarized below:[4]

3 Much of the information in this chapter is based on the personal observation of the writer while serving successively, during the period February 1942 to May 1944, in the following assignments: Flight Officer, Naval Air Station, Lakehurst, N.J.; Commander Airship Squadron Fourteen; Training Officer, Naval Airship Training Command.

4 Official Records, Naval Airship Training Command.

	ENTERED	COMPLETED	FAILED	PERCENT FAILURES
Officers	345	311	34	9.9
Aviation Cadets	1185	1083	102	8.6
Aviation Pilots	129	108	21	16.2
TOTAL	1659	1502	157	9.5

Of the 1083 aviation cadets who successfully completed the training, 125 were trained at Moffett Field. The remaining 958 either underwent all of their training at Lakehurst or received their advanced training there after the formation of the Naval Airship Training Command.

Thus, for the purposes of this study, three categories of candidates were available: Officers, aviation cadets, and enlisted men. Not only was the aviation cadet group the most populous of the three groups, but it was by far the most homogeneous group. Cadets were of the same age group -- between 18 and 26 years at the time of application. With few exceptions, the cadets had no previous naval experience. They were all of the same military rank, and most important, they had all taken the aptitude tests, which was not true in the case of officers and enlisted candidates. For these reasons the sample was limited to aviation cadets.

Inasmuch as this is a study of records which were not originally intended for the use attempted here, the availability of records also influenced the selection of the sample. Training and other records at Lakehurst were available to the writer whereas records at Moffett Field were not readily available. Consequently, the sample

was further delimited to those cadets whose entire training was accomplished at Lakehurst or who completed their training there. Other factors, to be discussed later also limited the sample.

The principal criterion against which the success of any selection program should be measured is success on the job. In the Navy this of course means the success of the individual officer in the performance of his assigned duties. When the definition of success is one of skill or technical competence in a restricted area, quantitative measures of performance are readily available. We merely need to measure the time it takes an individual to perform a certain task, or to measure his output in a given time. However when the definition of success contains many different aspects such as technical competence, leadership ability, resourcefulness, and many others, objective measures of performance are generally unavailable or unsuitable. It is necessary then to depend on estimates of performance such as subjective judgments or ratings made by observers who are familiar with the work of the persons rated.

In September 1943, the Commander Fleet Airwings Atlantic directed the Commanding Officers of all Airship Squadrons under his command to render ratings on all pilots on their "value as a pilot." These ratings were to be made on a five point scale and indicated as "either superior, above average, average, below average or unsatisfactory." Ratings of 637 pilots who had undergone airship pilot training as aviation cadets at Lakehurst were made available from the files for

this study.[5] Thus the sample was further limited to those cadets for whom specific ratings by commanding officers on their "Value as a pilot" were available.

It is realized that subjective ratings of an individual's performance of duty have certain limitations and may be subject to bias. Commanding officers frequently entertain misconceptions concerning the importance of particular characteristics of behavior.[6] Their judgments frequently reflect their likes and prejudices rather than variations in the performance of the individuals. The military rank of the officer being rated, his length of service in the particular organization and his total service in the Navy are illustrations of the variables which may cause a rater to err in his evaluation of an individual's level of competence. Despite these possible extraneous variables this criterion stands up well in that it has what may be called "face validity."[7] The ratings were made on the actual performance for which the subjects had been selected and trained, and ratings by commanding officers are used regularly in the naval service in evaluating personnel.

As a rough check on the reliability of the ratings, they were tabulated for each squadron and the number of ratings falling in

5 Historical Files, Commander Fleet Airships, Atlantic.

6 Guit, Dewey B. et al. Personnel Research in the Bureau of Naval Personnel. Princeton: Princeton University Press. 1948, p. 365.

7 Ibid. p. 392.

each category of the scale was determined. These tabulations revealed that except for one squadron the ratings were fairly well distributed, but were somewhat skewed in that there were many more in the "superior" than in the "unsatisfactory" category. This finding may be rationalized when one considers that most of the "unsatisfactory" category should have been eliminated in training. The anomalous squadron previously mentioned was badly skewed. In this squadron 33 percent were rated superior, 50 percent above average, and 16 percent average, with none in the below average or unsatisfactory categories. However this finding could likewise be rationalized. The squadron in question had been formed about six months prior to the time the ratings were made and was a specially selected group, chosen to make the trans-Atlantic flight and operate in the Mediterranean Theatre.

Another method of determining the validity of a selection device is by the success of the selectees in the training program. This may be based upon graduation or failure, or by grades and class standing in ground school. Success in school and success in an operating squadron usually are not so closely related that a measure of one may be substituted for a measure of the other. However when one considers that a person who fails a training program will not usually be allowed to perform the duties for which the program is the official prerequisite, then the criterion of success in school has definite practical validity.

A list of all cadets who failed the training course with the reasons for their failure, plus the rank order standings in ground school of the successful cadets were obtained from the records at Lakehurst. It was planned to use graduation or elimination from training as one of the criteria in this study. Unfortunately records of the flight aptitude test scores, which were maintained in the Navy Department, were available for only a small percentage of the "failure" group. There were so few individuals in this group whose test scores could be located that use of the pass/fail criterion was out of the question. Consequently rank order standing in the ground school was used as a criterion instead.

These standings were derived from grades based on the cadet's performance on non-standardized tests covering the subject matter of the various ground school courses.[8] These tests were prepared and administered by ground school instructors, most of whom were not skilled or trained in test construction. Grades assigned for each course were weighted, then averaged, to obtain the "final grade." From this final grade a rank order standing within his class was determined for each cadet. As Adkins points out: "The use of orders of rank is not generally a satisfactory method of portraying level of performance unless the total number of ranks

8 The general content of the airship pilot ground school course is included as Appendix II.

is also known."[9] A person ranking tenth in a group of fifteen is relatively low, whereas, in a group of fifty, a ranking of ten is relatively high. In order to equate the class standings for classes of different sizes, the rank order standings were transmuted to normalized standard scores. This consisted essentially of converting the ranks into "percent position" by means of the following formula:

$$\text{Percent Position} = \frac{100\ R - .5}{N}$$

where R is the rank of the individual in the series, and N is the number of individuals ranked. Then by referring these percent positions to a table for "The Transmutation of Orders of Merit into Units of Amount or 'Scores'", the individual's equated score on a ten point scale (0-9) was obtained.[10]

As noted previously, the test scores to be validated by these two criteria were obtained from the Navy Department. Again complete records for all the individuals were not available. Test scores on all these tests plus the F.A.M. index were obtained for 345 cadets (Sample A). In addition M.C.T. and A.C.T. scores plus F.A.M. were obtained for 341 cadets (Sample B). These two groups thus became the samples for this study. It is realized that there are many

9 Edwins, Dorothy C. et al. Construction and Analysis of Achievement Tests. Washington: U. S. Govt. Printing Office 1947, p. 138.

10 Garrett, Henry E. Statistics in Psychology and Education. New York: Longmans, Green and Co., 1947, pp. 174-175.

shortcomings to a sample arrived at in this manner. However, it was felt that in dealing with such a small universe plus the limitations imposed by lack of test results, this was the only method of securing samples of suitable size. Inquiry was made in the Navy Department to determine whether the availability or non-availability of test scores was the result of any bias relevant to the purposes of this study. There was no evidence to show that any such bias existed. Despite the method of obtaining them, there is evidence to indicate that these samples are not as tainted as one might suspect. This is discussed in the following chapter.

The correlations made in this study were based on distributions with few class-intervals, in order to approximate the results to be expected with the five letter-grades used in the tests and F.A.R. Inasmuch as administrative decisions concerning the selection of cadets for entry into the training program involved letter-grades not raw scores, it is appropriate that the usefulness of the tests be evaluated in terms of comparably coarse groupings.

In addition to the tests and criteria two other items which might have some bearing on the overall relationship between the test scores and performance were included in the correlations. These were "education" and "length of service." Education was divided into three categories: (1) Non-High School Graduates; (2) High School Graduates with less than two years college; and (3) Cadets with two or more years of college.

Length of service was the period of time, in months, from the date of completion of training to the date the commanding officers' ratings were made. This varied from four to nineteen months in Sample A, and from four to thirty months in Sample B.

Finally, there is always the danger that extraneous factors may contribute to the final results of any study by introducing irrelevant sources of variation. One of these variables which might have been operating is motivation for airship pilot training. There was no separate recruitment for L.T.A. training. Initially cadets for airship training were selected from some stage of L.T.A. flight elimination and frequently the cadets who had already been eliminated were likely contenders for this group. Beginning in June 1942 the first stage of training for all cadets was the Navy Pre-Flight School. At this stage cadets were informed of the L.T.A. program and given an opportunity to volunteer for it. While it is undoubtedly true that certain individuals had entered the cadet program in order to obtain L.T.A. training most cadets had little previous information concerning airships. It soon became common knowledge however that the L.T.A. pilot training period was much shorter than H.T.A. The expectation of a shorter period before obtaining wings, increased salary, rank, commission, rather than a sincere interest and desire to fly airships, may have influenced some cadets to volunteer for airship training.

Another type of extraneous variable is the chance assignment of flight instructors, and the attendant non-uniform quality of instruction. A poor or unsympathetic instructor, or one whose personality is incompatible with the cadets' can have considerable influence on the final performance. The reverse of course is also true.

A final type of extraneous factor is the chance duty assignment upon successful completion of the training course. It may be that pilots assigned to a well-organized and established squadron have fewer opportunities for learning all aspects of their duties and for demonstrating their abilities, than if they had been assigned to a more recently formed squadron. The net result of this inequality is that they will be evaluated on quite different standards. If the above extraneous factors are operating the criterion measure will reflect the differences in performance, but the source of the differences can not be traced to the usual predictive factors.

CHAPTER IV

ANALYSIS OF DATA

Three pencil-and-paper tests, the A.C.T., M.C.T., and E.I. were administered to 1185 aviation cadets who eventually entered airship pilot training. Of the 1083 cadets who successfully completed training, 958 received all or part of their training at Lakehurst. Subjective ratings by commanding officers on their "Value as a pilot" were made on 637 members of the latter group while they were serving as pilots in a fleet airship squadron. Complete records of test scores and the F.A.R. index were available for 249 of the 637 cadets, and all scores except E.I. were available for 341 cadets. These two groups formed the samples for this study.

In evaluating the predictive value of these tests, the commanding officer's rating and rank order standing in the ground school were used as criteria. In addition the effect of previous education and length of service was evaluated. Data was scaled as follows:

1. Class Rank — recorded as normalized standard scores on a 0—9 scale.
2. Commanding Officer's Rating — recorded in terms of a 1—5 rating.
3. Flight Aptitude Rating — recorded in terms of a 1—5 rating.
4. Aviation Classification Test — recorded in terms of a 1—4 rating.

5. Mechanical Comprehension Test — recorded in terms of a 1—4 rating.

6. Biographical Inventory — recorded in terms of a 1—5 rating.

7. Education — recorded in terms of a 0—2 rating.

8. Length of Service — recorded in terms of a 4—10 rating.

Based on the above data, Pearsonian product-moment intercorrelations among the three tests, the F.A.R., and Education were computed. Test scores, the F.A.R. index, and length of service were correlated with the two criteria by means of product-moment coefficients. Tables 1(A) and 1(B) show the resultant intercorrelations and correlations, (A) and (B) referring to Sample A and Sample B respectively.

TABLE I

Intercorrelation of Tests, and Correlation With Criteria

Sample A - All Variables N = 249

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	M	S.D.
Class Rank(1)	-.05	.19*	.01	-.08	.00	.16**	.06	4.61	1.96
F.A.R. (2)		.10	.79*	.79*	-.31*	-.08	-.13**	2.83	1.35
A.C.T. (3)			.96*	-.07	.03	.03	-.22*	2.56	.72
M.C.T. (4)				.21*	-.11*	-.08	-.05*	2.59	.73
B.I. (5)					-.19*	-.03	.00	3.00	1.00
Education (6)						.01	-.02	.91	.77
C.O.Rating(7)							.28*	2.85	.65
Service (8)								7.09	2.93

* - Significant at the .01 level
** - Significant at the .05 level

Sample B (BI Missing) N = 341

	(2)	(3)	(4)	(6)	(7)	(8)	M	S.D.
Class Rank(1)	-.03	.15*	.08	.07	.15*	.02	4.60	1.89
F.A.R. (2)		.10	.64*	-.18*	-.08	-.11	2.77	1.21
A.C.T. (3)			.38*	.21*	.14**	-.02	2.31	.82
M.C.T. (4)				-.03	.00	-.08	2.46	.77
Education (6)					.08	.12**	1.06	.76
C.O.Rating(7)						.34*	2.74	.72
Service (8)							8.42	3.94

* - Significant at the .01 level.
** - Significant at the .05 level.

The intercorrelations among the three tests show the greatest relationship between the A.C.T. and the M.C.T. This relationship may be in part due the common factor of reasoning ability in both tests. The correlation between B.I. and M.C.T. may be partly explained by the association between ability and interest. Individuals with the ability to understand certain physical laws and principles may tend to develop interests in those activities which require such abilities. The slightly negative correlation between the B.I. and the A.C.T. is to be expected for these predictors measure completely different aspects of the individual.

In connection with the intercorrelations among the tests it is of interest to note that they have a high degree of consistency with those presented in another study.[1]

In Table II are presented the intercorrelati ns between B.I., A.C.T., and M.C.T. The first three columns are based on three different samples as indicated in the C.A.A. Report. The N's in

1 Civil Aeronautics Authority Division of Research. Report No. 20. "The History and Development of the Biographical Inventory" p. 19.

each of these three samples were greater than 1000 cases. The fourth and fifth columns are Samples A and B respectively from Table I of this study. The consistency between these two groups of intercorrelations leads one to believe that the samples used in this study may be representative despite the method by which they were obtained.

TABLE II

Intercorrelations Among M.I., A.C.T., and M.C.T.

	C.A.A. Report				
	Jul-Nov '41	Mar-Apr '42	Sept-Oct '42	: A	B
A.C.T. and M.I.	.04	.05	.01	:-.07	—
A.C.T. and M.C.T.	.33	.30	.29	: .38	.36
M.I. and M.C.T.	.23	.25	.20	: .21	—

The high correlation between the T.A.M. and both the M.I. and M.C.T. is explained by the fact that the T.A.M. is an index derived from the M.C.T. and M.I. scores. In effect it is a built-in correlation. The low degree of relationship between the A.C.T. and T.A.M. may be in part due to the curtailment of range in A.C.T. scores by the requirement of a minimum score on that test.

At first glance one may question the relationship between A.C.T. and "Education". In one sample there is practically no relationship and in the second, although significant, the correlation is quite low. When we consider that the A.C.T. is a discriminatory measure of the general intelligence of the individual, whereas "Education" may be considered a coarse index of the amount of knowledge to which he has been subjected, then this lack of relationship is more meaning-

ful. Granted that academic progress is a somewhat selective procedure, unless there is some reference to the rate of progress or to the grades received, academic status is not necessarily a reflection of intellectual capacity. "Education" as categorized here may be as much an indication of economic status as of intellectual capacity.

Although a significant correlation between length of service and the commanding officer's rating was expected, its magnitude is not sufficient to affect this study. Inspection of Table I will show that the validities of the tests are such that they will not be changed materially by partialling out the effect of length of service from the criterion measure (C.O. Rating).

Table I indicates that in both samples the correlation between the three tests and the class rank criterion are significant only in the case of the A.C.T. The slightly negative relationship of the S.I. with class rank is not unusual for the Inventory is not designed to measure this factor. The correlation of A.C.T. and class rank is lower than one might expect. Two reasons may be advanced for this low correlation and for the almost total lack of relationship between M.C.T. and class rank. First is the extraneous factor of motivation. While the attainment of a satisfactory grade in ground school was necessary for successful completion of the training program, the course of study was not considered too formidable. Certain individuals were undoubtedly satisfied to merely pass the examinations with a minimum expenditure of effort. There was of

course no way to measure this variable. The second reason is concerned with the criterion itself. As stated previously, the examinations upon which class standings were based were usually non-standardized and of unknown reliability and validity. Tests of this sort quite frequently measure a verbal factor more than the knowledge and skills which they purpose to measure. For this reason they may improperly discriminate against the individual who has mastered the content of the course but lacks the verbal facility to express it properly.

Further inspection of Table 1 reveals that except for the A.M.T. in the sample there is no relationship between the tests and performance on the job as expressed by the commanding officer's rating. There are two basic inferences which we may deduce from this finding. The first is that the tests do not adequately discriminate among the candidates. When one considers that the tests were not designed for use in the selection of airship pilots this is a logical deduction. The second deduction is that commanding officers' ratings are not a true measure of performance in this case. Admittedly there are many limitations to the use of ratings by superiors as a criterion. The "Value as a pilot" was not defined in any way, nor was the rating analytical or designed to provide internal checks. However, as pointed out previously, the ratings were made on the actual performance for which the subjects had been trained, by officers who constantly used this system of evaluating personnel. Thus, in this case, the criterion has a believable face validity.

CHAPTER V

SUMMARY AND CONCLUSIONS

Three pencil-and-paper tests were used to screen applicants for airship pilot training during World War II: the Aviation Classification Test (A.C.T.), the Mechanical Comprehension Test (M.C.T.), and the Biographical Inventory (B.I.). The A.C.T. is a general intelligence test developed specifically for Naval Aviation and designed to eliminate those candidates who would be too dull to follow complicated orders or profit from technical ground training. Groups of items in this test deal with Judgment, Number Facility, Motor Reading and Comparisons. The M.C.T. consists of items dealing with pictured mechanical situations in which the verbal factor is minimized, and is more a measure of the candidate's ability to handle the mechanical concepts of everyday life. The items of the B.I. deal with bio-mechanical topics, interest, habits, attitudes, and preferences of the individual. Keys for this test were developed by item analysis of response patterns of cadets who passed or failed heavier-than-air (H.T.A.) flight training.

These three tests were originally developed as an aid in the selection of candidates for H.T.A. flight training. With the expansion of lighter-than-air (L.T.A.) flight training in 1942, and due to the peculiarities of the service, the same tests were used in the selection of candidates for this branch of Naval Aviation, despite certain fundamental differences between the two types of flying. The tests were originally standardized in terms of the

relative percentages of men failing in flight training at each score level. Continuous surveys of the effectiveness of these tests have been maintained, but always in terms of success or failure of individuals in the U.T.A. flight training program. These tests are still being used to screen candidates for U.T.A. pilot training, without ever having been validated in this field. This study is the first attempt to evaluate their predictive value in terms of individuals in the U.T.A. training program.

This report, based upon records which were not originally intended for such use, was limited by the availability of data concerning test scores. As a result, the subjects in the two samples reported on were confined to those individuals for whom complete records were available. Correlations between each test and two criteria, obtained for each of the two samples, are given. The criteria used were the class standing of the cadets in ground school and a subsequent rating by the individual's commanding officer on his "value as a pilot". In addition intercorrelations between the three tests, the candidate's previous education, and his length of service at the time of evaluation by his commanding officer are given. While the correlation of the M.C.T. and I.I. with both criteria was negligible, the A.C.T. predicted success in ground school to a slight degree. Due to this lack of overall correlation between tests and criteria, the intercorrelations

among the tests and the other variables are of little value in the sense that they are not able to suggest combinations of scores to better predict the criterion.

From an analysis of these data either or both of two possible conclusions may be reached:

1. The criteria used in this study do not discriminate adequately among the candidates. That is, they are not true measures of potential performance.

2. The tests, designed for N.C.A. selection, do not discriminate adequately between good and poor airship pilots.

Certain limitations or inadequacies in class rank as a criteria were recognized. First, it was not a total measure of success in the training program, but rather one of the two major factors upon which the successful completion of training was based. Secondly, the validity of the objectives and course content of the ground school was never systematically evaluated in terms of their relationship to the actual operational job to be performed. Finally, the validity of the grades and marks were open to question. As stated previously, they were largely subjective and were probably influenced in varying degrees by instructor biases and prejudices as well as by chance error.

The inadequacy of ratings as a criterion measure has also been discussed previously. However, it is desirable to emphasize that the ratings used in this study were made within a single category and not on several different traits. Also, they were made by Officers

who regularly used this system of evaluation and were therefore presumably qualified to make the desired discriminations. Furthermore, the ratings were made on the actual performance for which the subjects had been trained.

If we grant that either or both of the criteria are discriminatory, if only to a moderate degree, then the second conclusion is tenable. In other words, the tests are not valid predictors of success in airship pilot training. This, however, does not necessarily mean that none of the items of the tests apply, but rather that the number of items dealing with success as airship pilots is small or that the items, especially in the B.I., are weighted wrong in scoring. This indicates the need for an item analysis of the A.C.T. and M.C.T. to determine discriminatory items in the airship field. In addition, the development of a separate scoring key for the B.I., based on an analysis of response patterns of successful airship pilots may prove fruitful.

Either of the two conclusions may be correct or, more likely, both of these factors have probably operated to reduce the validity coefficients. In any event, regardless of which conclusion we accept, the fact still remains that there is no evidence to show that the three tests now being used in the selection of candidates for airship pilot training are valid as predictors of success in this field. In other words, the administrative procedure used for selection of this group may serve to exclude as many potentially

successful candidates as it includes.

This leads to two suggestions for further study. These suggestions should be carried out concurrently. One is that an investigation be made of the reliability and, to whatever extent possible, the validity of various criteria in A.T.C. training. The other is that experiments be conducted to develop new tests designed on the basis of job analysis of the airship pilot's duties. Solution of these two problems will insure the most effective use of the candidates available for entrance to airship pilot training.

AIR SPEED

ALTITUDE

THE AVIATION CLASSIFICATION TEST

SAMPLE PROBLEMS

Instrument Reading, or Ability to Follow Directions.

In these questions you will be asked to do one or more of the following operations:

CT	Close Throttle	IP	Increase Propeller Pitch	RS	Raise Stabilizer
OT	Open Throttle	DP	Decrease Propeller Pitch	LS	Lower Stabilizer

To perform an operation, black out the space or spaces under the proper symbols to the right of the question. The first question has been marked correctly. Look at the meters at the top of the page.

1. If the air speed is less than 155 mph, open the throttle.

1.	CT	IP	RS
	▮ OT	DP	LS

2. If the airplane is above 1500 feet, decrease the propeller pitch, if not above 1500 feet, lower the stabilizer.

2.	CT	IP	RS
	OT	▮ DP	LS

Vocabulary.

1. Awkward means the opposite of

- 1—strong
- 2—pretty
- 3—short
- ▮ 4—graceful
- 5—swift

Comparisons.

Indicate whether the two groups of letters and figures are same (s) or different (d).

s	d	
	▮	1—PBy 64 *** PBy—64
▮		2—486397 *** 486397
▮		3—NC404179 *** NC404179
	▮	4—SOS2pde5,1256***SOSqde5,1256

Practical Judgment.

1. A destroyer convoying four merchant ships off her own coast hears at dawn an SOS from a small pleasure boat from a position 30 miles off her course. Her skipper should:

- 1—Proceed at once to the rescue.
- 2—Change the course of the convoy toward the ship sending the SOS.
- ▮ 3—Disregard the SOS and proceed on her course.
- 4—Break radio silence and call the nearest anti-submarine patrol headquarters for orders.
- 5—Ask the advice of the captains of the merchant ships.

Arithmetic.

1. 14.8564 rounded to the nearest tenth is

- 1—15.0
- 2—14.86
- ▮ 3—14.9
- 4—14.85
- 5—14.0

2. As a naval aviator, which of these problems would you not try to solve by means of plane trigonometry?

- 1—The height of a cloud base.
- 2—The distance across a lake.
- 3—Your altitude above the ground.
- ▮ 4—The capacity of your gas tank.
- 5—The length of a strange runway.

580432 O - 44

Ground School Curriculum -- Airship Pilot Training

GROUP I

AERODYNAMICS

AEROSTATICS

AIRMANSHIP

AIRSHIP MAINTENANCE AND DOCK ROUTINE

MOORED MASTLING AND MOORING

LIFTING GASES

MATERIALS, DESIGN AND CONSTRUCTION

STRATEGY, TACTICS AND MISSION

GROUP II

METEOROLOGY

COMMUNICATIONS

LIFE SAVING EQUIPMENT

NAVIGATION

ORDNANCE AND GUNNERY

PHOTOGRAPHY

POWER PLANTS

R. O. TRAINING

The subjects listed in Group I are concerned only with airship operations. Those in Group II are common to both L.T.A. and H.T.A. operations and the treatment is fundamentally the same.

APPENDIX II

BIBLIOGRAPHY

Books

Adkins, Dorothy C. et al. Construction and Analysis of Achievement Tests. Washington: U. S. Government Printing Office, 1947.

Allen, Hugh. The Story of the Airship (Non-rigid). Chicago: The Lakeside Press, 1942.

Burtt, Harold Ernest. Principles of Employment Psychology. New York: Harper and Bros., 1942.

Dubois, Philip H. et al. Army Air Forces Aviation Psychology Research Reports — Report No. 2. Washington: U. S. Government Printing Office, 1947.

Flanagan, J. C. et al. Army Air Forces Aviation Psychology Research Reports — Report No. 1. Washington: U. S. Government Printing Office, 1948.

Garrett, Henry E. Statistics in Psychology and Education. New York: Longmans, Greene, and Co., 1947.

Ghiselli, Edwin E. and Brown, Clarence W. Personnel and Industrial Psychology. New York: McGraw Hill, 1948.

Mosher, W. E. and Kingsley, J. D. Public Personnel Administration. New York: Harper and Bros. 1941.

Munsterberg, H. Psychology and Industrial Efficiency. Boston: Houghton Mifflin, 1913.

Scott, W. D. et al. Personnel Management: New York: McGraw Hill Book Co. Inc., 1941.

Stuit, Dewey B. et al. Personnel Research and Test Development in the Bureau of Naval Personnel. Princeton, N. J.: Princeton University Press, 1947.

Tiffin, Joseph. Industrial Psychology. New York: Prentice Hall Inc., 1947.

Viteles, Morris S. Industrial Psychology. New York: W.W.Norton and Co. Inc., 1932.

Periodicals

Fiske, Donald W. "Validation of Naval Aviation Cadet Selection Tests Against Combat Criteria." J. Appl. Psych. Vol. 31, No. 6, December 1947.

Jenkins, John G. "Naval Aviation Psychology." Psychol. Bull. 1946, 43; Amer. Psychol. 1946, 1.

_____. "Prediction of Flight Training Performance by Biographical Data." Journal of Aviation Medicine, 1944, 15.

McComas, H. C. "Time Tests for Efficiency in Telephone Operation." J. Appl. Psych. and Scientific Method 11, 1914.

Myers, C. W. "Psychological Tests for Pilots and Stenographers." J. Appl. Psych. 1, 1917.

Rosendahl, C. W. "Report on the Lighter-than-air Situation" made before the National Aeronautics Committee of the American Legion. 24 August 1947.

Scott, W. D. "Scientific Selection of Salesmen." Advertising and Selling. 25, 1915.

Official Publications

"Construction of Keys A, B, and C for the Biographical Inventory." Navy Department, Aviation Psychology Technical Memorandum No. 3. 1 Nov. 1944.

"Notes on Psychology and Personality Studies in Aviation Medicine." Medical Manual for Air Department. Washington: U. S. Government Printing Office. January 1941.

"Psychology for Naval Leaders." Naval Leadership -- Book II. Washington: U. S. Government Printing Office. 1942.

"The History and Development of the Biographical Inventory". U.S.N. Division of Research, Report No. 7". Washington: October 1946.

"The Predictive Value of Naval Aviation Cadet Selection Tests." Navy Department Naval Aviation Letter. Vol. 2, No. 4, March 1947.

Viteles, Morris S. et al. "The Aircraft Pilot -- 5 Years of Research" National Research Council. Washington: June 15 1945.

DATE DUE			